Human-AI Collaboration in the Fashion Design Process

Andrea Tocchetti[1], Giulia Monterosso[1], Francesca Palazzetti Romualdi[1], Paola Bertola[2], Marco Brambilla[1], Angelica Vandi[2], Federica Vacca[2]

[1]Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Via Giuseppe Ponzio, 34, Milano, 20133, IT
[2]Politecnico di Milano, Dipartimento di Design, Via Giovanni Durando, 10, Milano, 20158, IT
{firstname.lastname}@polimi.it

Abstract

The recent development of Generative AI (GenAI) revolutionized the fashion industry, automating tasks like market analysis and trend forecasting, as well as innovating cloth design. However, its potential application in supporting designers in the fashion design process for creativity purposes has yet to be explored. Our research studied the complexity and pitfalls of such a process through interviews with domain experts. Starting from that assessment, this article compares generic and context-specific GenAI tools from several perspectives to explore their features and capabilities in enhancing the cooperation between fashion designers and AI tools and agents in the creative process. The most promising systems are tested on typical creative tasks, including generating images from textual descriptions (specified according to technical criteria), generating variations of existing designs and pictures, and generating photographic images from sketches. The experiments involved two data sets and five domain-specific tasks. The effectiveness and limitations of the tools are evaluated quantitatively to assess how they can properly support the designers' work.

Introduction

Recent improvements in Artificial Intelligence (AI) have driven a technological revolution in several fields, promoting human-AI collaboration to achieve high performance and efficiency. Such developments also concern the fashion industry, in which Generative AI (GenAI) technologies are increasingly being applied, bringing novel dynamics and shaping clothes' conceptualization, creation, and manufacturing (Shi, Wong, and Zou 2025). In fashion design, *i.e.*, a process requiring creativity, patience, accuracy, time, and precision, AI can support designers in designing new clothes or customizing collections to meet customers' requests (Guo et al. 2023) and even forecast future trends through data and behaviour analyses (Imtiaz et al. 2024). Furthermore, GenAI can improve the clothes' design process, which involves several laborious steps, *i.e.*, inspiration, sketching, picking materials, and prototyping. This article explores potential integrations of AI in fashion processes to enhance human creativity and operational efficiency, finally driving AI-designer cooperation. Interviews with experts identified key domain challenges; generic and context-specific GenAI solutions were analyzed starting from this assessment. A final expert-in-the-loop assessment with real data was performed to analyze each implementation's capability to provide innovative and creative designs.

Figure 1: The basic steps of the fashion design process.

Background

Generative AI is relatively new in fashion design and is mainly applied to process management rather than creativity. Despite its novelty, its impact is groundbreaking and covers various aspects from virtual try-on technologies and hyper-personalization (*e.g.*, providing tailored shopping tips and supporting product discovery) (Gou et al. 2023) to on-demand manufacturing and supply chain management (Imtiaz et al. 2024). Furthermore, AI-driven market analysis allows fashion companies to predict trends, adapt to the market's demands, and remain competitive (Imtiaz et al. 2024).

The most interesting aspect in which GenAI can be effective is design generation, *i.e.*, supporting designers in thinking out of the box and creating unique competitive designs (Guo et al. 2023) by translating sketches and mood boards into high-fidelity 3D models and providing suggestions through data analytics. Additionally, commonly used design tools can be replaced by AI in the early stage of fashion design, increasing productivity and allowing for automated drawing generation (Shi, Wong, and Zou 2025). Despite such potential changes, designer-AI cooperation is still central and subject to scrutiny and research. A known case is a collaboration between the AWAYTOMARS clothing community and IBM Watson that allowed the creation of novel designs (Moreau, Prandelli, and Schreier 2023).

Methodology

A traditional design process includes several complex steps: initial research on current trends, detailed sketching and conceptualization, sketch refinement and technical drawing cre-

ation, and dress production (see Figure 1). Design experts from the Gianfranco Ferré Research Centre in Milan were interviewed to understand such a process, its problems and requirements, reporting on its importance concerning quality and creativity and highlighting its intricacy because of potential imprecision and sketching uncertainty, as well as its elevated time requirements. Such problems slow the creative process, limiting the designer's potential for innovating and responding to changing market trends. Consequently, a more efficient and less costly design process is needed while achieving high quality and creativity.

In the considered process, AI was proven to contribute by automating some preliminary steps, *e.g.*, analyzing market trends and making the process less time-consuming, or quickly generating realistic pictures of models wearing dresses from textual descriptions or sketches drawn by fashion designers. Such cooperation greatly enhances the creative process, providing high flexibility by allowing quick changes and, consequently, facilitating prototyping.

To foster adoption and provide an accurate assessment of the effectiveness of these approaches, fashion companies and researchers need to evaluate the available AI tools and their application to the design process by analyzing their features and testing their performance (*i.e.*, usability, efficiency, crash frequency, and customization). This work considers text-to-image (starting from free text and structured text), image-to-image (to create variations of clothes), image-to-sketch, and sketch-to-image as fundamental tasks. These are implemented in a hybrid human-AI team of Generative AI agents and human fashion designers. Additionally, the study considers the capability of the tools to be integrated into broader ecosystems (*i.e.*, through API and customer support) and grow with future needs (*i.e.*, scalability and availability).

In this work, we report the results of a validation experience of the most promising tools tested in real scenarios on the aforementioned tasks using two datasets, each including fashion images and their metadata (*i.e.*, detailed descriptions, materials used, and stylistic annotations). Data was cleaned and structured to ensure a proper analysis. Each scenario was assessed based on the fidelity of the output to the input(s), the potential presence of errors in the generated picture, and the proper modification of the input image (only in image-to-image tasks).

AI Tools' Analysis

This study considers four context-agnostic AI models (*i.e.*, ChatGPT, Microsoft Copilot, DreamStudio, and Leonardo AI) and a context-specific one (*i.e.*, The New Black AI, an AI system directly addressing fashion designers' needs), and analyzes their features to identify the best one(s) to support designers. In particular, this preliminary analysis evaluated each system's usability, reliability, customization, functionality, scalability, APIs, and customer support.

- *Usability*. ChatGPT and Copilot are easier to use, accepting simple garment descriptions as input. Instead, Leonardo AI and DreamStudio require additional parameters (*e.g.*, the required image style and a score representing the fidelity to the initial prompt) to extend textual prompts, leading to lower intuitiveness. Furthermore, it was tested that the latter tools might require multiple prompts to achieve acceptable outcomes.
- *Reliability*. Compared to the other context-agnostic technologies, ChatGPT encountered the least problems and generation errors (*e.g.*, DreamStudio might occasionally generate a blurred image). Furthermore, it generated images with higher quality and efficiency under several aspects (*e.g.*, level of detail). Additionally, all generic GenAI tools besides ChatGPT required higher computational time when overly detailed prompts were provided.
- *Customization*. Only ChatGPT and Leonardo AI allow for custom data uploading and model training, leading to more accurate outputs in context-specific scenarios. Furthermore, only ChatGPT allowed custom training on specific tasks.
- *Functionality*. All generic AI tools can perform text-to-image, image-to-image, and its derived tasks, *i.e.*, image-to-sketch and sketch-to-image, while only ChatGPT and Copilot support image-to-text and sketch-to-text.

The New Black AI was proven intuitive, easy to use thanks to the guidelines and explanations provided, and capable of quickly generating detailed images. Although it generates high-quality pictures, a few problems might arise when providing complex and overly detailed prompts. The tool offers tailored functionalities for fashion designers, *e.g.*, the possibility to upload a dataset for sketching or designing purposes, further enhancing AI-designer cooperation. The New Black AI supports text-to-image, image-to-image, image-to-text, and their derived tasks, achieving better image generation and description than context-agnostic approaches.

Additionally, all systems provide APIs that support tools' integration into software components. Furthermore, each system offers customer support through various options, *e.g.*, live chat, call centre, forums, and documentation. Regarding systems' scalability, they are all highly scalable due to their advanced infrastructure, *e.g.*, Kubernetes (ChatGPT), cloud computing (The New Black AI), etc. Finally, all tools can be accessed via their web interface, although only ChatGPT and Copilot provide mobile and desktop applications.

Considering the analyzed aspects, ChatGPT and The New Black AI were identified as the most promising systems for AI-designer cooperation, as they provide (among others) higher-quality content, ease of use, and customization.

Experiments & Discussion

The considered AI tools were tested on two datasets across five different tasks, *i.e.*, free-text text-to-image, guided text-to-image (*i.e.* a text-to-image task in which the prompt is structured as a list of features instead of a free-text, as shown in Figure 2), image-to-image, image-to-sketch, and sketch-to-image. The considered datasets involved a selection of images from the spring/summer 2004 prêt-à-porter collection from the Gianfranco Ferré Research Centre by the artist Vittorio Zecchin (referred to as the Ferré dataset) and a set of web-sourced images from well-known fashion

brands. Each image included a photo of (a model wearing) the dress and a list of metadata (*e.g.*, title, location, detailed dress description, date, and more). Datasets were cleaned and organized to fit the tasks, resulting in 450 images for the Ferré dataset and 40 images for the web-sourced one. A total of 350 tests were performed, 200 on the first dataset and 150 on the latter. The results of such analyses covered the fidelity towards the input prompt (referred to as fidelity), towards the input image (referred to as accuracy), and the generation errors. AI's outcomes were assessed by fashion design experts from the Gianfranco Ferré Research Centre through 5-point and 7-point Likert-scale questions. Several aspects were evaluated for each dress and task to determine the final fidelity and accuracy scores. In particular, the dress type, features (*i.e.*, proper representation of the distinctive characteristics of the garment), usage (*i.e.*, proper representation of the context in which the dress is worn), style, colour, and fabric (*i.e.*, whether the proper fabric was used) were assessed for the guided text-to-image task. In contrast, clothing (*i.e.*, whether the garment was properly generated), colour, textiles (*i.e.*, whether the proper textiles were used), and details (*i.e.*, whether the generated image includes the details required) were assessed for all the other tasks.

Text-to-Image. The models' capability to translate textual prompts (input) into images (output) was assessed. The considered models achieved good fidelity in the free-text text-to-image task, reporting about a 78% score in clothing and colour and 65% in textiles. The New Black AI underperformed slightly in details, reporting an average of 57% compared to 66% from ChatGPT. Regarding generation errors, both ChatGPT and The New Black AI made very few mistakes, mainly related to human anatomy and garments, respectively. Experiments on the Ferré dataset reported lower textiles and details scores than the web-sourced dataset, with an average decrease of 9% and 8%, respectively. When guided prompts are used, an even higher fidelity is attained. Both models performed well in type, usage, and colour, reporting an average of 80%, 77.5%, and 79%, respectively. In all other aspects, The New Black AI outperformed ChatGPT, reporting an average of 72% and 56% across all scores, respectively. In particular, both models mainly encountered difficulties in generating dresses with good fabric fidelity, reporting an average of 63% for The New Black AI and 46% for ChatGPT. Very few generation errors were encountered when guided prompts were used. Experiments on the Ferré dataset reported lower scores than the web-sourced dataset in features, usage, and fabric, with an average decrease of 14%, 10%, and 5%, respectively. The type and colour scores were constant, while the fabric score increased by 8%.

Image-to-Image. The models' capability to apply image changes through textual prompts (input) was assessed. In particular, this task covered the models' capability to add or remove details to garments. ChatGPT underperformed on average compared to The New Black AI in all aspects, with a fidelity score of 62.5% and 69.5%, respectively. Very few errors were identified for either model. Additionally,

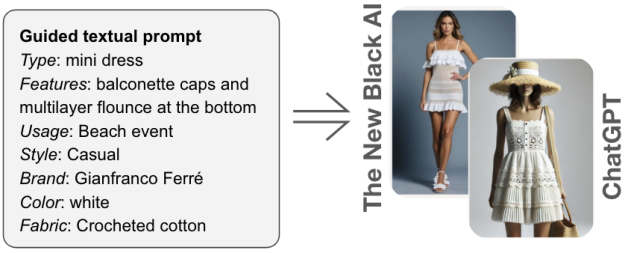

Figure 2: An example of a text-to-image task with a guided prompt. In particular, the prompt is organized as a list of features rather than a free-text describing the dress.

Figure 3: An example of an image-to-image task, starting from an image of a model wearing a dress and a prompt describing how to modify the dress.

The New Black AI demonstrated outstanding performances in updating the original image instead of generating a different one (see Figure 3). Experiments on the Ferré dataset reported lower scores than the web-sourced dataset in textiles and details, with an average decrease of 20% and 28%, respectively.

Image-to-Sketch. The model's capability to generate garment sketches (output) from images (input) was assessed, demonstrating The New Black AI's very scarce capability to generate such content. For example, in Figure 4, the AI only applies a black-and-white filter instead of producing the sketch of the given image. For this reason, the model was excluded from the analysis for this task. On the other hand, ChatGPT achieved a 36% accuracy, demonstrating there's still room for improvement. Experiments on the Ferré dataset reported an increase in accuracy of 26% compared to the web-sourced dataset.

Sketch-to-Image. The model's capability to generate images (output) from garment sketches (input) was assessed. This task was assessed only on the Ferré dataset as the web-sourced one contains no sketch. Like the previous task, The New Black AI failed to properly understand the prompt, generating images with unnecessary elements (*e.g.*, the designer's annotations reported on the dress' sketch). Instead, ChatGPT demonstrated good performance (see Figure 5), reporting an average fidelity of 66%. Furthermore, the model could not capture all the details, generating images with errors mainly due to the high im-

Task	Measure	ChatGPT	The New Black AI
(Free-Text) Text-to-Image	Fidelity	72.5%	70%
(Guided) Text-to-Image	Fidelity	67.8%	73.7%
Image-to-Image	Fidelity	62.5%	69.5%
Image-to-Sketch	Accuracy	36%	-
Sketch-to-Image	Fidelity	66%	-
	Accuracy	32%	-

Table 1: A table representing ChatGPT's and The New Black AI's average fidelity towards the input prompt and accuracy towards the input image across all tasks and datasets.

Figure 4: An example of an image-to-sketch task using The New Black AI. Instead of generating a sketch of the image, the AI applies a black-and-white filter.

Figure 5: An example of a sketch-to-image task using ChatGPT. In this instance, the tool generates the dress accurately, considering even the small details in the sketch.

age complexity and resulting in an average accuracy of 32%.

In summary, both AI tools demonstrated their strengths in supporting designers and enhancing their creativity in the fashion design process (see Table 1). However, there's still room for improvement in some aspects (*e.g.*, image-to-sketch and sketch-to-image tasks). In particular, while The New Black AI demonstrated good fidelity and high consistency in text-to-image and image-to-image tasks, ChatGPT was proven effective and superior in fidelity when unstructured prompts were provided. Additionally, ChatGPT demonstrated its capability to perform additional tasks (*i.e.*, image-to-sketch and sketch-to-image), although it only achieved basic results. Variations in performance were identified based on the dataset. In particular, when the web-sourced dataset was used, better performances (with a few exceptions) were achieved, most likely due to the higher level of details of the images as they involve novel and innovative designs from well-known brands.

Conclusions

This article discussed the potential applications of AI in supporting fashion designers through AI-designer cooperation. AI tools were analyzed by inspecting their features and testing the most promising ones against five fashion-related tasks to assess their performance, thus highlighting their effectiveness and limitations. Notice that GenAI techniques could be applied to other stages of the fashion design pipeline and can be combined into novel tools, further mixing their strengths and enhancing designers' capabilities. In conclusion, while several legal and ethical concerns are yet to be addressed within the fashion community regarding the usage of AI, *e.g.*, sensitive subjects, gender discrimination, and intellectual property, AI tools represent groundbreaking technology capable of driving innovation.

References

Gou, J.; Sun, S.; Zhang, J.; Si, J.; Qian, C.; and Zhang, L. 2023. Taming the Power of Diffusion Models for High-Quality Virtual Try-On with Appearance Flow. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, 7599–7607. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701085.

Guo, Z.; Zhu, Z.; Li, Y.; Cao, S.; Chen, H.; and Wang, G. 2023. AI Assisted Fashion Design: A Review. *IEEE Access*, 11: 88403–88415.

Imtiaz, A.; Pathirana, N.; Saheel, S.; Karunanayaka, K.; and Trenado, C. 2024. A Review on the Influence of Deep Learning and Generative AI in the Fashion Industry. *Journal of Future Artificial Intelligence and Technologies*, 1(3): 201–216.

Moreau, P.; Prandelli, E.; and Schreier, M. 2023. Generative Artificial Intelligence and Design Co-Creation in Luxury New Product Development: The Power of Discarded Ideas. *SSRN Electronic Journal*.

Shi, W.; Wong, W.; and Zou, X. 2025. Generative AI in Fashion: Overview. *ACM Trans. Intell. Syst. Technol.*